

Extrait du Spyworld Actu

<http://www.spyworld-actu.com/spip.php?article2148>

# Le datamining en quête de démocratisation

- Informatique - Software -



Date de mise en ligne : lundi 19 juin 2006

---

**Spyworld Actu**

---

Alors qu'elles ont repris leurs investissements dans les outils de gestion de la relation clients et que le secteur de l'aide à la décision retrouve sa vigueur, les entreprises disposent d'une masse considérable d'informations sur leurs clients. Données que les outils d'aide à la décision ont bien du mal à transformer en opportunités commerciales.

Xavier Dugast, PDG de Spad, éditeur français spécialisé dans le datamining, ou exploitation de gros volumes de données, défend cette thèse : « Le marché est tiré par la croissance du volume de transfert des informations. Son développement est exponentiel et double le volume des données tous les dix mois. Plus on a de données à traiter, plus il est compliqué de trouver de l'information. Le décisionnel s'arrête au reporting ; le datamining va au-delà. »

Si, jusqu'à présent, les outils de datamining ont été adoptés par quelques secteurs économiques, vont-ils parvenir à se diffuser plus largement ?

« Les choses bougent, assure David Williamson, directeur des opérations de SPSS. Les applications de datamining ne sont plus uniquement mises en oeuvre par les services R&D des grands comptes, mais directement par le marketing. Il y a une volonté d'utiliser le datamining à des fins tactiques. Pour cela, il faut disposer d'outils simples à utiliser, qui permettent de produire un modèle dans un délai inférieur à la semaine. » Mais, si les logiciels de datamining se sont améliorés, s'il est moins nécessaire d'apprendre un langage aussi spécialisé que celui de SAS et si l'industrialisation dans la production et la réutilisation des modèles est à l'ordre du jour, la qualité des outils ne suffit pas. « Un bon traitement de texte ne suffit pas pour devenir écrivain, ironise Stéphane Tufféry, consultant statisticien en datamining. Même si l'outil de datamining peut être utilisé très simplement en déplaçant des icônes et des flèches sur un plan de travail graphique, les statistiques présentent de nombreux pièges, et il est très facile de mettre en évidence des corrélations qui n'existent pas. »

### **Bousculer la suprématie de SAS**

SPSS cherche à apporter des éléments de réponses avec des offres tout-en-un d'analyse prédictive issues du rachat du spécialiste DataDistilleries, fin 2003. « Les affaires sont plus petites mais en plus grand nombre, note David Williamson. Autre signe de la démocratisation : nous sommes de plus en plus sollicités par des partenaires du monde de l'aide à la décision. Nous avons des partenariats avec Micropole-Univers, Business & Décision et Oracle. » L'éditeur de Clementine joue sur la distribution pour étendre l'adoption de son offre et pour remettre en cause la domination de SAS.

Hervé Kauffmann, directeur de Kxen, parie, pour sa part, sur une rupture technologique pour parvenir au même but. « Actuellement, il faut plusieurs semaines pour développer un modèle statistique, explique-t-il. Plutôt que les théories de Fisher communément adoptées, Kxen a implémenté la théorie mathématique de Vladimir Vapnik. On a ainsi réduit le temps de traitement des données, assuré par un automate. Ces outils de troisième génération brisent la limite de la taille des fichiers que l'on étudie. Par exemple, Wanadoo travaille sur des segmentations de trois mille colonnes, alors que, avec SAS, il n'est possible de travailler que sur quinze à vingt colonnes. »

La volumétrie reste un vrai problème. Le Crédit Lyonnais, qui compte six millions de clients, a mis en place très tôt un entrepôt de données, comme le rappelle Guillaume de Beauvais, chef de produit de LCL : « Historiquement, LCL a été une des premières banques à créer un entrepôt de données sur lequel on a « pluggé » l'outil SAS. Cet entrepôt fait à peu près 4 To. On brasse énormément de données : contacts commerciaux et informations comptables. On a une volumétrie de données énorme et on est obligé de raboter dans tous les coins. Il y a six ans, nous avions un historique de deux ans ; il est maintenant réduit à un an et demi. » Si la technologie de Kxen répond effectivement à la problématique de la volumétrie, elle n'est pas applicable partout.

### Des outils pour tous ?

L'élargissement de la cible des solutions de datamining est-elle d'actualité ? Hervé Kauffmann en est persuadé : « Potentiellement, tous les utilisateurs de progiciels SAP et Siebel vont avoir besoin d'outils de datamining. Tout un chacun devra être capable de les utiliser. Le futur est en partie là, et Kxen a été conçu pour ça. » Néanmoins, beaucoup reste à faire pour convaincre les entreprises d'investir dans des outils onéreux et complexes à mettre en oeuvre, même pour celles qui exploitent au quotidien une base de données marketing.

Le site de location de DVD Glowria.fr stocke ainsi, depuis novembre 2003, l'ensemble des données de ses vingt-deux mille clients, soit deux millions d'informations. « On possède un catalogue de dix mille titres et les gens connaissent mieux les blockbusters et les séries TV que tout le reste, explique Nathalie Jacquot, responsable du recrutement on line. L'objectif est de faire tourner le catalogue pour conserver le client et, ainsi, faire du chiffre d'affaires. » Pour segmenter ses clients, le site s'appuie sur des outils internes et sur les briques Olap et Reporting Services de SQL Server.

Même constat chez Assurland, site de comparaison d'assurances, qui gère une base de un million et demi d'internautes. Stanislas Di Vittorio, fondateur et président du directoire d'Assurland, détaille son architecture technique : « La base de production, qui fonctionne sur un cluster SQL Server 2000, est purgée chaque nuit dans notre entrepôt de données. Nous nous sommes dotés de la plate-forme BusinessObjects Web Intelligence, qui fait l'historique de l'ensemble de nos données depuis l'origine de la société. La mise en place de Business Objects a représenté un investissement important et, désormais, toutes nos équipes sont allouées au développement des nouveaux produits. Nous connaissons bien le monde de l'assurance, et je ne suis pas persuadé que l'impact financier apporté par les outils de type SAS sur notre activité serait considérable. »

## Si vous êtes pressé

- ▶ Dominé par SAS et SPSS, le marché du datamining reste un domaine de spécialistes, sur quelques secteurs économiques.
- ▶ La multiplication des entrepôts de données dans les entreprises, bâtis sur les données des progiciels pour les outils d'aide à la décision, constitue une opportunité pour les éditeurs de datamining.
- ▶ Les solutions restent chères à mettre en place et nécessitent des compétences pointues en statistiques.

## Un marché français encore très ouvert

La dernière étude menée par IDC auprès de deux cents entreprises françaises dresse l'état du marché français des solutions de gestion de la relation clients. 85 % des entreprises disposent désormais d'une base de données clients. Elles sont en revanche moins du tiers à avoir mis en place des modèles prédictifs et à pratiquer le scoring comportemental de leurs clients. Une aubaine pour les éditeurs, qui vont pouvoir diffuser leurs solutions de datamining en dehors des banques, assureurs et opérateurs télécoms, leurs principaux utilisateurs.

Dans la dernière édition de son carré magique, le Gartner classe SAS et SPSS comme leaders du marché, largement devant Fair Isaac, Kxen, Chordiant Software et Portrait Software/Quadstone. Teradata est toujours

considéré comme un acteur de niche.

## Ils ont dit

« Un retour sur investissement très facile à démontrer »

Stéphane Tufféry, consultant statisticien en datamining, auteur de Datamining et statistique décisionnelle, aux Éditions Technip

La conjoncture n'incite pas forcément les entreprises moyennes à investir dans les technologies de datamining, mais il n'y a pas de raison pour que ce marché ne se développe pas. On peut facilement démontrer la valeur d'un outil de datamining par un calcul d'appétence, par exemple. Un test en aveugle sur des commerciaux va vite démontrer la valeur du ciblage via ce type d'outil par rapport à un ciblage traditionnel. C'est facile et objectif, il suffit alors de comparer les taux de retour.

*Post-scriptum :*

<http://www.01net.com/article/319621.html>